CLAIMS AS AMENDED

1. A method for optimal multimedia content delivery over networks from a server to [client] one or more clients, comprising [the steps of]:

[D] delineating a state variable that represents the data rate to each client;

[D] delineating a set of [requirements] conditions which represent the time-varying constraints on the data rate of said multimedia content[, given by] said conditions including:

(1) [T] the total data rate for all clients does not exceed the maximum throughput of the server or network, whichever is least;

(2) [T] the data rate from server to client does not exceed the maximum data rate for the client;

(3) [T] the data rate of the client will never overflow the client buffer;

(4) [T] the server will never underflow; and

[(6)](5)[T] the data rate from the server will never be less than the client's minimum data rate, which is a non-increasing function of time obtained by dividing the content not yet delivered by the remaining play time;

[D] delineating a cost function which represents the value of a proposed solution; and

[P] performing periodic computations [to solve said inequalities] in compliance with conditions (1) - (5) to obtain [the] a state value that maximizes said cost function.

2. A method as in claim 1, [further comprising] wherein said conditions further include

(6) [T] the current maximum client data rate is given by the minimum of:

[T] the stored initial maximum client data rate;

[T] the data rate required to fill the remaining client buffer during the current of said periodic computations;

[T] the data rate required to complete the delivery of said multimedia content;

[T] the client data rate never exceeds said current maximum client data rate.

whereby said current maximum client data rate is periodically recomputed to maintain an optimal solution over a given period of time.

3. A method as in claim 2, [further comprising] wherein:

[S] said cost function represents maximal throughput and is given by the sum of said client data rates for all active clients.

4. A method as in claim 2, [further comprising] wherein:

[S] said cost function represents maximal charge and is given by the sum for all active clients of[:] said client data rates times the client's cost of service[.].

5. A method as in claim 3 for bandwidth allocation for delivery of multimedia data from server to one or more clients over a network, comprising the steps of:

[D] determining the maximum flow rate and minimum flow rate for each client;

[D] determining the flow rate range for each client as given by the difference between said maximum flow rate and said minimum flow rate;

[I] initializing current flow rate for each client as said minimum flow rate and summing said flow rate into total server flow rate; and

[A] allocating remaining server bandwidth to remaining clients until[l] they each saturate or no bandwidth remains.

6. A method as in claim 5 wherein said step of allocating remaining server bandwidth to remaining clients [is done fairly by a procedure that comprises the steps of] comprising:

[S] sorting the list of clients according to said flow rate range; [and]

[D] determining equally-allocated remaining server bandwidth if allocated evenly to all remaining unprocessed clients [and];

[D] determining the range of remaining client bandwidth as given by the difference between said maximum flow rate and said minimum flow rate; and

[D] determining saturation by comparing said equally-allocated remaining server bandwidth and said range of remaining client bandwidth, and allocating the lesser of these two amounts to each remaining client flow rate;

[W] whereby allocating flow to remaining clients based upon the sorted client range flow rates and determining allocation of remaining server bandwidth based upon a

comparison of saturation of server versus saturation of each client maximizes allocation of total bandwidth for maximal flow rate to maximum number of clients.

7. A method as in claim 4 for bandwidth allocation for delivery of multimedia data from server to one or more clients over a network, comprising the steps of:

[D] determining the maximum flow rate and minimum flow rate for each client;

[D] determining the flow rate range for each client as given by the difference between said maximum flow rate and said minimum flow rate;

[S] sorting the list of clients according to said flow rate range;

[I] initializing current flow rate for each client as said minimum flow rate and summing said flow rate into total server flow rate; and

[A] allocating remaining server bandwidth to remaining clients such that lower paying clients receive bandwidth only if higher paying ones are saturated.

8. A method as in claim 7 wherein said step of allocating remaining server bandwidth to remaining clients comprises the steps of:

[F] for each remaining unprocessed client:

[D] determining equally-allocated remaining server bandwidth if allocated evenly to all remaining unprocessed clients [and] ;

[D] determining the range of remaining client bandwidth as given by the difference between said maximum flow rate and said minimum flow rate;

[D] determining saturation by comparing said equally-allocated remaining server bandwidth and said range of remaining client bandwidth, and allocating the lesser of these two amounts to each remaining client flow rate; and

[W] whereby allocating flow to remaining clients based upon the sorted client range flow rates and determining allocation of remaining server bandwidth based upon a comparison of saturation of server versus saturation of each client maximizes allocation of total bandwidth for maximal flow rate to maximum number of clients.

9. A method for connection acceptance control for delivery of multimedia data from server to one or more clients over a network, comprising the steps of:

[D] determining [the] server swing capacity given by the difference between the total server bandwidth and the sum of the minimum flow rates of all currently-connected clients [rate for each client]; and

[A] allocating server bandwidth for each prospective client which will fit without server bandwidth saturation, as determined by comparing [the] an average data play rate of each prospective client with the remaining bandwidth, represented by said server swing capacity, available to the server,

wherein the minimum flow rate for each client is expressed as a non-increasing function of time obtained by dividing content not yet delivered by remaining play time.

10. A method as in claim 9 wherein said remaining bandwidth available to the server is given by said server swing capacity.

11. A method as in claim 10 wherein said remaining bandwidth available to the server is give by said server swing capacity less a server flow safety margin, thereby allowing server capacity to be subsequently lowered by up to the safety margin without requiring load shedding, and without affecting client sessions in process.

12. A method as in claim 9 wherein said step of allocating server bandwidth for each prospective client which will fit without server bandwidth saturation [is further comprised of] comprises:

[A] allocating server bandwidth to each prospective client sequentially until a prospective client is located in which said average data play rate exceeds said server swing capacity.

13. A method as in claim 9 wherein said step of allocating server bandwidth for each client which will fit without server bandwidth saturation [is further comprised of] comprises:

[A] allocating server bandwidth to each prospective client sequentially for each client which can be activated without server bandwidth saturation.

14. A method for bandwidth allocation for delivery of multimedia data from server to one or more clients over a network, comprising [the steps of]:

[S] storing a sequence of data representing scheduled bandwidth changes for the server;

[D] determining the maximum flow rate and minimum flow rate for each client at the present time, the determination of the minimum flow rate being based on a non-increasing function of time obtained by dividing content not yet delivered by remaining play time;

[D] determining the flow rate range for each client as given by the difference between said maximum flow rate and said minimum flow rate;

[S] sorting the list of clients according to said flow rate range;

[I] initializing current flow rate for each client as said minimum flow rate and summing said flow rate into total server flow rate; and

[A] allocating remaining server bandwidth to remaining clients.